

WHAT IS CLAIMED IS:

1 1. A method for indexing data, comprising:
2 receiving a token;
3 determining whether a data field associated with the token is a fixed width;
4 when the data field is a fixed width, designating the token as one for which fixed
5 width sort is to be performed; and
6 when the data field is a variable length, designating the token as one for which a
7 variable width sort is to be performed.

1 2. The method of claim 1, wherein the token is variable width and further
2 comprising:
3 transforming the variable width token into a fixed width token.

1 3. The method of claim 1, further comprising:
2 performing a fixed width sort on one of dual code paths and a variable width sort
3 on the other of dual code paths.

1 4. The method of claim 1, further comprising:
2 generating a sort key that includes a token type, a token, a document identifier, a
3 document section, and an offset in a document.

1 5. The method of claim 1, further comprising:
2 receiving different sections of a document at different times.

1 6. The method of claim 5, wherein the different sections include a context
2 section and an anchor text section.

1 7. The method of claim 1, further comprising:
2 generating sort keys for each token of multiple tokens; and

3 using the sort keys to create posting lists that simultaneously are ordered by token
4 and by document identifier for each token.

1 8. The method of claim 7, further comprising:
2 using the sort keys to bring together multiple sections of a document.

1 9. The method of claim 1, further comprising:
2 sorting on certain bits of a sort key containing multiple bits.

1 10. The method of claim 9, further comprising:
2 sorting on uppermost bits of the sort key.

1 11. A computer system including logic for indexing data, comprising:
2 receiving a token;
3 determining whether a data field associated with the token is a fixed width;
4 when the data field is a fixed width, designating the token as one for which fixed
5 width sort is to be performed; and
6 when the data field is a variable length, designating the token as one for which a
7 variable width sort is to be performed.

1 12. The computer system of claim 11, wherein the token is variable width and
2 wherein the logic further comprises:
3 transforming the variable width token into a fixed width token.

1 13. The computer system of claim 11, wherein the logic further comprises:
2 performing a fixed width sort on one of dual code paths and a variable width sort
3 on the other of dual code paths.

1 14. The computer system of claim 11, wherein the logic further comprises:

2 generating a sort key that includes a token type, a token, a document identifier, a
3 document section, and an offset in a document.

1 15. The computer system of claim 11, wherein the logic further comprises:
2 receiving different sections of a document at different times.

1 16. The computer system of claim 15, wherein the different sections include a
2 context section and an anchor text section.

1 17. The computer system of claim 11, wherein the logic further comprises:
2 generating sort keys for each token of multiple tokens; and
3 using the sort keys to create posting lists that simultaneously are ordered by token
4 and by document identifier for each token.

1 18. The computer system of claim 17, wherein the logic further comprises:
2 using the sort keys to bring together multiple sections of a document.

1 19. The computer system of claim 11, wherein the logic further comprises:
2 sorting on certain bits of a sort key containing multiple bits.

1 20. The computer system of claim 19, wherein the logic further comprises:
2 sorting on uppermost bits of the sort key.

1 21. An article of manufacture including a program for indexing data, wherein
2 the program causes operations to be performed, the operations comprising:
3 receiving a token;
4 determining whether a data field associated with the token is a fixed width;
5 when the data field is a fixed width, designating the token as one for which fixed
6 width sort is to be performed; and

7 when the data field is a variable length, designating the token as one for which a
8 variable width sort is to be performed.

1 22. The article of manufacture of claim 21, wherein the token is variable
2 width and wherein the operations further comprise:
3 transforming the variable width token into a fixed width token.

1 23. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 performing a fixed width sort on one of dual code paths and a variable width sort
4 on the other of dual code paths.

1 24. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 generating a sort key that includes a token type, a token, a document identifier, a
4 document section, and an offset in a document.

1 25. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 receiving different sections of a document at different times.

1 26. The article of manufacture of claim 25, wherein the different sections
2 include a context section and an anchor text section.

1 27. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 generating sort keys for each token of multiple tokens; and
4 using the sort keys to create posting lists that simultaneously are ordered by token
5 and by document identifier for each token.

1 28. The article of manufacture of claim 27, wherein the operations further
2 comprise:
3 using the sort keys to bring together multiple sections of a document.

1 29. The article of manufacture of claim 21, wherein the operations further
2 comprise:
3 sorting on certain bits of a sort key containing multiple bits.

1 30. The article of manufacture of claim 29, wherein the operations further
2 comprise:
3 sorting on uppermost bits of the sort key.